Molecular Plant • Volume 2 • Number 5 • Pages 933–942 • September 2009 RESEARCH ARTICLE

Identification of Lignin and Polysaccharide Modifications in *Populus* Wood by Chemometric Analysis of 2D NMR Spectra from Dissolved Cell Walls

Mattias Hedenström[a,1], Susanne Wiklund-Lindström[a], Tommy Öman[a], Fachuang Lu[b], Lorenz Gerber[c], Paul Schatz[d], Björn Sundberg[c] and **John Ralph[b,d]**

a Department of Chemistry, Umeå University, SE-90187 Umeå, Sweden
b Department of Biochemistry, and Great Lakes Bioenergy Research Center, University of Wisconsin, Madison, WI 53706, USA
c Umeå Plant Science Centre, Department of Forest Genetics and Plant Physiology, Swedish University of Agricultural Sciences, SE-90183 Umeå, Sweden
d US Dairy Forage Research Center, USDA-Agricultural Research Service, Madison, WI 53706, USA

ABSTRACT 2D ^{13}C–^{1}H HSQC NMR spectroscopy of acetylated cell walls in solution gives a detailed fingerprint that can be used to assess the chemical composition of the complete wall without extensive degradation. We demonstrate how multivariate analysis of such spectra can be used to visualize cell wall changes between sample types as high-resolution 2D NMR loading spectra. Changes in composition and structure for both lignin and polysaccharides can subsequently be interpreted on a molecular level. The multivariate approach alleviates problems associated with peak picking of overlapping peaks, and it allows the deduction of the relative importance of each peak for sample discrimination. As a first proof of concept, we compare *Populus* tension wood to normal wood. All well established differences in cellulose, hemicellulose, and lignin compositions between these wood types were readily detected, confirming the reliability of the multivariate approach. In a second example, wood from transgenic *Populus* modified in their degree of pectin methylesterification was compared to that of wild-type trees. We show that differences in both lignin and polysaccharide composition that are difficult to detect with traditional spectral analysis and that could not be *a priori* predicted were revealed by the multivariate approach. 2D NMR of dissolved cell wall samples combined with multivariate analysis constitutes a novel approach in cell wall analysis and provides a new tool that will benefit cell wall research.

Key words: Aspen; biostatistics; cell walls; multivariate data analysis; NMR spectroscopy; tension wood.

INTRODUCTION

Plant cell walls provide massive amounts of polysaccharides and lignin, important for a variety of industrial processes. Research aimed at optimizing cell wall characteristics for end users by conventional or molecular breeding is highly dependent on reliable methods to determine cell wall chemistry and composition in novel genotypes. This is a challenging task because each wall component alone is complex and the native polymer composition and chemical structure of cell walls remain difficult to elucidate (Boerjan et al., 2003; O'sullivan, 1997). Furthermore, the interactions between wall polymers are only poorly characterized and understood. The most common approach in secondary cell wall analysis has been to isolate and purify specific components, and to characterize them in detail using a variety of degradative and spectroscopic analytical tools. Wet chemistry approaches suffer from unavoidable destruction of molecular structure. Isolations are often accompanied by unwanted modifications, and unfortunate partitioning due to the incomplete extractability of the component of interest. Moreover, polymer interactions in the cell wall will remain undetected.

Ideally, the chemical structure of the cell wall and its components should be resolved in intact samples containing the native cell wall. Several methods are in place for this purpose,

[1] To whom correspondence should be addressed. E-mail: mattias.hedenstrom@chem.umu.se, fax (+46)907867655, tel. (+46)907865788

doi: 10.1093/mp/ssp047, Advance Access publication 12 July 2009
Received 29 April 2009; accepted 16 June 2009

such as solid-state NMR, NIR, FT-IR and Raman spectroscopy. While these methods are rapid (with the exception of solid-state NMR), they provide limited information on chemical structure, which make them more useful as tools for broad fingerprinting or chemical imaging (Gierlinger and Schwanninger, 2006; Maunu, 2002; Tsuchikawa, 2007). A promising approach to obtain more detailed information on native composition and chemical structure of cell walls is the dissolution of whole cell wall samples combined with high-resolution multidimensional NMR analysis. Dissolution followed by acetylation provides acetylated cell walls that are soluble in common NMR solvents. High-resolution 2D ^{13}C–^{1}H correlative NMR methods allowed, for example, minor lignin components to be examined in the presence of overwhelming polysaccharides (Lu and Ralph, 2003; Ralph and Lu, 2004). The dissolution solvents used, N-methylimidazole (NMI) or tetrabutylammonium fluoride in dimethylsulphoxide (DMSO), appear to be non-destructive as determined using lignin model substrates (Lu and Ralph, 2003). However, successful dissolution requires extensive cleaving of the polymers into smaller fragments via ball-milling that unavoidably will introduce some destruction of the native wall structure. Studies suggest that ball-milling cleaves the weakest bonds (glycosidic linkages in polysaccharides, β-aryl ethers in lignins), leaving most of the structural features intact (Hu et al., 2006; Ikeda et al., 2002). A further improvement to cell wall dissolution is the use of NMI-d_6 that allows direct NMR analysis of the cell walls without acetylation (Yelle et al., 2008). The method provides some increase in the dispersion of polysaccharide anomeric peaks and allows the natural acetylation of wall mannans and xylans to be investigated. Simply swelling cell walls in DMSO-d_6, alone, produces a gel that also allows the use of solution-state NMR (Kim et al., 2008). The spectra have lower resolution than for a true solution (due to the viscosity and other factors), but spectra can be acquired more rapidly and are surprisingly rich in resolved detail.

A 2D ^{13}C–^{1}H HSQC NMR spectrum of the dissolved wall provides a wealth of information and represents a high-resolution compositional and structural 'fingerprint' of all wall components, where many of the peaks have been assigned to chemical structures (Kim et al., 2008; Lu and Ralph, 2003; Ralph et al., 1999; Ralph et al., 2005; Yelle et al., 2008). To date, however, full use of this information has been limited by the extent to which the spectra can be analyzed. Traditionally, this is done by peak picking and area measurements of well resolved peaks from known structures supposed to differ between samples. This approach is limited by *a priori* hypotheses about annotated structures, and peak picking of overlapping peaks is notoriously difficult. The complete spectral information is thus not used, and structural differences in the cell wall between samples might therefore pass undetected. Alternatively, spectra can be analyzed through an unsupervised approach by creating differential spectra between sample types. Whereas this approach uses the full spectral information, it does not provide a comprehensive overview of the sample set that is important, for example for detection of outliers. Furthermore, the relative importance of different peaks for sample discrimination is difficult to assess.

Multivariate data analysis can be used to take full advantage of the information present in 2D NMR spectra of dissolved cell walls. Principal component analysis (PCA) (Jackson, 1991; Wold et al., 1987) and orthogonal projections to latent structures (OPLS) (Trygg and Wold, 2002) have proven to be tremendously useful in the analysis of complex spectral data (Holmes et al., 2000; Stenlund et al., 2008; Wold et al., 1997). These methods can be used to build statistical models where all peaks in the spectra, assigned or not, are considered simultaneously. PCA is a projection method that gives a comprehensive overview of systematic variation within complex data, where clustering of samples with similar features as well as eventual outliers can be detected. PCA is therefore commonly used as a first step in multivariate data analysis. OPLS, on the other hand, is a supervised regression method that relates the variation within the data to a class identifier, such as a specific genotype. Differences between sample types that are obscured by systematic biological or experimental variation in a PCA can thus still be identified using OPLS. When applied to a two-class problem, this is called OPLS-DA (DA for discriminant analysis). The use of multivariate analysis of 2D NMR spectra as a global analysis technique to study the chemical structure and composition of cell walls is thus highly justified. It allows the visualization of differences between samples as highly informative 2D loading spectra, as recently demonstrated in a study of extracted pectin fractions (Hedenström et al., 2008). These 2D loading spectra can be interpreted in the same way as the original spectra and thus provide a detailed picture of structural differences in the cell wall between sample types.

In this paper, we demonstrate the power of combining 2D NMR spectra from dissolved whole cell wall materials with multivariate analysis to reveal structural/compositional differences between wood samples from *Populus.* In the first case, tension wood (TW) is compared to normal wood (NW). This provides an example in which the major compositional differences are well known (Timell, 1969), and can be considered as a positive control for the multivariate approach. In the second case, we compared a transgenic *Populus* line down-regulated in the activity of a pectin methylesterase (PME) with wild-type trees. This provides an example in which differences in cell wall composition are subtle and in which no prior predictions about modifications to major wall components were possible. In both cases, multivariate analysis of 2D NMR spectra revealed differences beyond those readily detected by traditional interpretation of NMR spectra. These two examples demonstrate that the approach can be used to effectively and reliably reveal structural/compositional differences in all components of the plant cell walls induced by environmental or genetic modifications. An advantage that cannot be overstated is that the major differences between sample types described by the multivariate models can be visualized as 2D loading spectra that allow substantial interpretation of those differences.

Moreover, the 2D loading spectra can be used to highlight the peaks with the highest significance for sample type discrimination. We propose that this technology has the potential of becoming an important tool in research aimed at optimizing cell wall characteristics, through breeding or genetic approaches.

RESULTS AND DISCUSSION

A 2D ^{13}C–^{1}H HSQC NMR spectrum from a dissolved and acetylated cell wall sample from *Populus* wood consists of a large number of peaks, originating from correlations between carbons and their directly attached protons, where the composition of both polysaccharides and lignin components are observable (Figure 1). Many peaks are well resolved but there are also regions in the spectra where a number of peaks overlap (mostly from different polysaccharides). Peak assignment is, of course, crucial for interpretation of these spectra and, while the assignment of the different lignin structures has been verified with model substances, several of the polysaccharide assignments shown in Figure 1 still remain tentative and some sugar units, such as rhamnose and fucose, are not assigned in those spectra. Complete assignment of those cell wall components requires further studies. 2D ^{13}C–^{1}H HSQC spectra have previously been used to elucidate lignin structure in cell walls (Ralph et al., 2006, 2008; Wagner et al., 2007), but, in this paper, we also demonstrate the potential for a more comprehensive analysis including all cell wall components.

Comparison of Cell Wall Composition between *Populus* Tension Wood and Normal Wood

As a first example to demonstrate the reliability of combining 2D NMR analysis of dissolved cell walls with multivariate

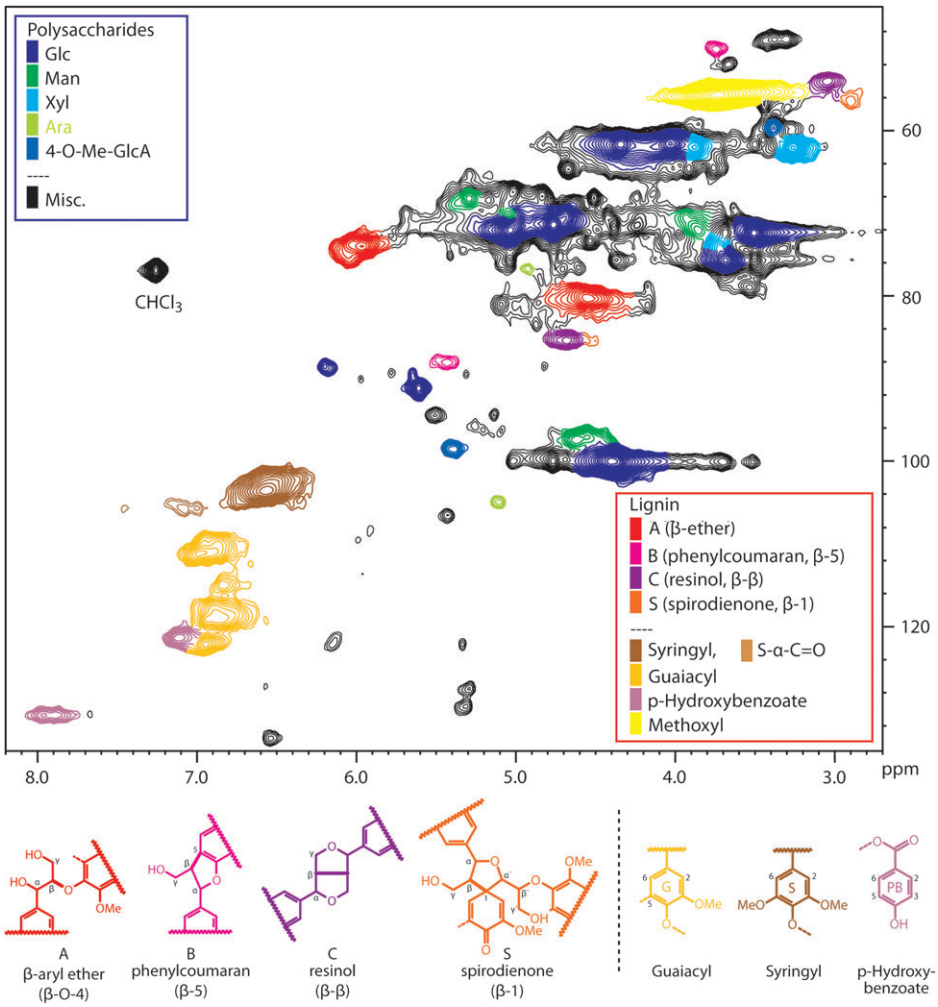

Figure 1. Typical 2D ^{13}C–^{1}H HSQC Spectrum of Dissolved Acetylated Cell Wall from *Populus*.
^{1}H and ^{13}C chemical shifts are located on the *x*- and *y*-axis, respectively. The large number of resolved (and unresolved) peaks reflects the complexity of chemical structures of the cell wall components. Peak annotations are based on assignment of model substances and literature values. Acetate peaks that appear at ~2/30 ppm in the $^{1}H/^{13}C$ dimensions are not shown. The polysaccharide abbreviations refer to the following residues in their respective polymers: Glc, Glucosyl; Man, mannosyl; Xyl, xylosyl; Ara, arabinosyl; 4-OMe-GlcA, 4-O-methyl glucuronic acid.

analysis to detect differences in cell wall composition between sample types, we compared tension wood (TW) with normal wood (NW) from aspen (*Populus tremula* L.) The most prominent feature of TW is the presence of an inner cellulose-rich wall layer (the so-called G-layer) in the wood fibers, which results in an increase in wood cellulose and a corresponding decrease in lignin, xylan, and glucomannan (Timell, 1969). TW lignin is also known to have an increased syringyl/guaiacyl (S/G) ratio (Pilate et al., 2004; Yoshida et al., 2002) and small increases in galactan, arabinogalactan proteins (AGP), and rhamnogalacturonan I (RG I) (Bowling and Vaughn, 2008; Timell, 1969).

Multivariate data analysis on full-resolution spectra, without the need for peak picking or integration of peak volumes before analysis, was accomplished according to Hedenström et al. (2008) (Figure 2). Data pre-treatment (as described in Methods) effectively removed unwanted peaks, from residual solvent, or data points containing only noise. After normalization to a constant sum and mean centering, the resulting spectral data were subjected to PCA to obtain an overview of the variation between spectra and hence also the difference in wood composition between sample types. The fact that a clear separation between NW and TW samples is seen along the first principal component ($\mathbf{t}_1$) indicates a significant difference in wood composition between these sample types (Figure 3A). The majority (94%) of the variation within the data was explained by the two calculated principal components (81% in the first and 13% in the second). The loading vector along the first principal component thus contains spectral information highly correlated to this sample discrimination and was subsequently back-transformed to a 2D loading spectrum (Figure 3D). In this spectrum, the value of each variable (spectral data point) corresponds to its correlation with $\mathbf{t}_1$. Positive peaks are thus correlated with samples with positive score values along $\mathbf{t}_1$ in Figure 3A, namely TW samples. Comparison with previously made assignments (Figure 1) verified that peaks positively correlated with the TW samples originated mainly from cellulose, reflecting the higher cellulose content in TW. The increase in cellulose was accompanied by a relative reduction in lignin, xylan, and mannan (from glucomannan). In addition, a few unassigned peaks were also elevated in tension wood. These unidentified peaks found to co-vary with cellulose were initially suspected to be remnants from material not fully acetylated in the sample preparation but this possibility could be ruled out, as they later were found to be unaffected by acetylation conditions. Furthermore, the appearance of these peaks in cellulose reference spectra implies that they originate either from cellulose itself or from a hemicellulose tightly associated with cellulose. Delineating their exact nature will require further investigation.

It can be noted that the NW samples are more tightly clustered along both $\mathbf{t}_1$ and $\mathbf{t}_2$ than the tension wood samples (Figure 3A). The loadings $\mathbf{p}_2$ (not shown) that describe the spectral variation along $\mathbf{t}_2$ indicate that this difference results from a phasing error associated with the cellulose peaks, which explains why it is more pronounced in the cellulose-rich TW samples. The ability to investigate intra- as well as inter-class variation to detect sources of variation induced by, for example, experimental error is an appealing feature of this multivariate approach. In principle, variation in any direction in the score plot can be studied.

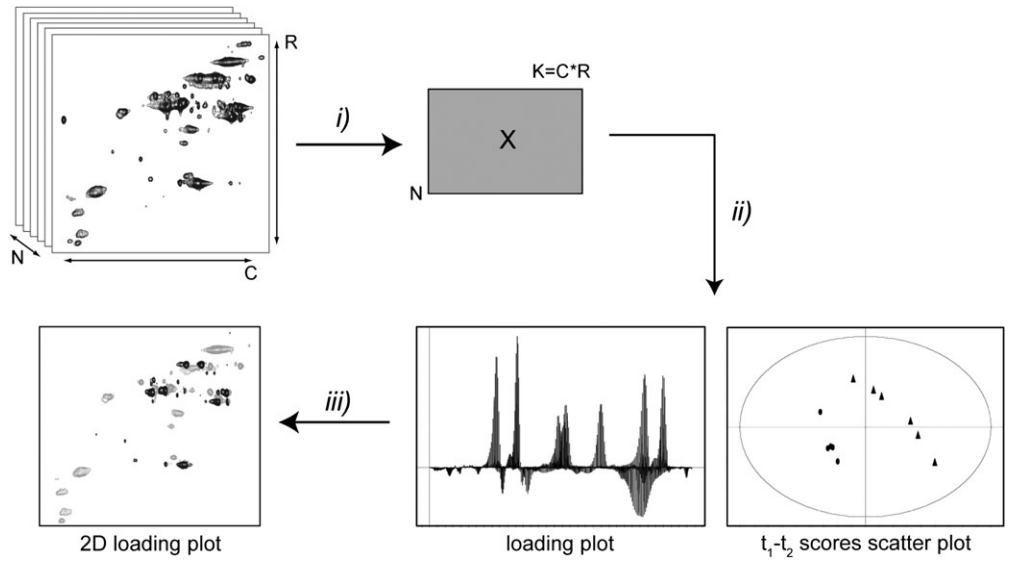

Figure 2. Overview of the Procedure for Using Multivariate Analysis on 2D NMR Data.

(A) Each spectrum is converted to a row vector and placed in a new data matrix **X.** Data points with intensity below a set threshold were considered noise and removed from the matrix.

(B) Scores and loadings resulting from multivariate analysis of matrix **X** are analyzed in order to detect latent structures within the data.

(C) The loadings, initially represented as line plots of length K (number of columns in **X**), are converted to 2D loading spectra by reversing the unfolding procedure described in (A).

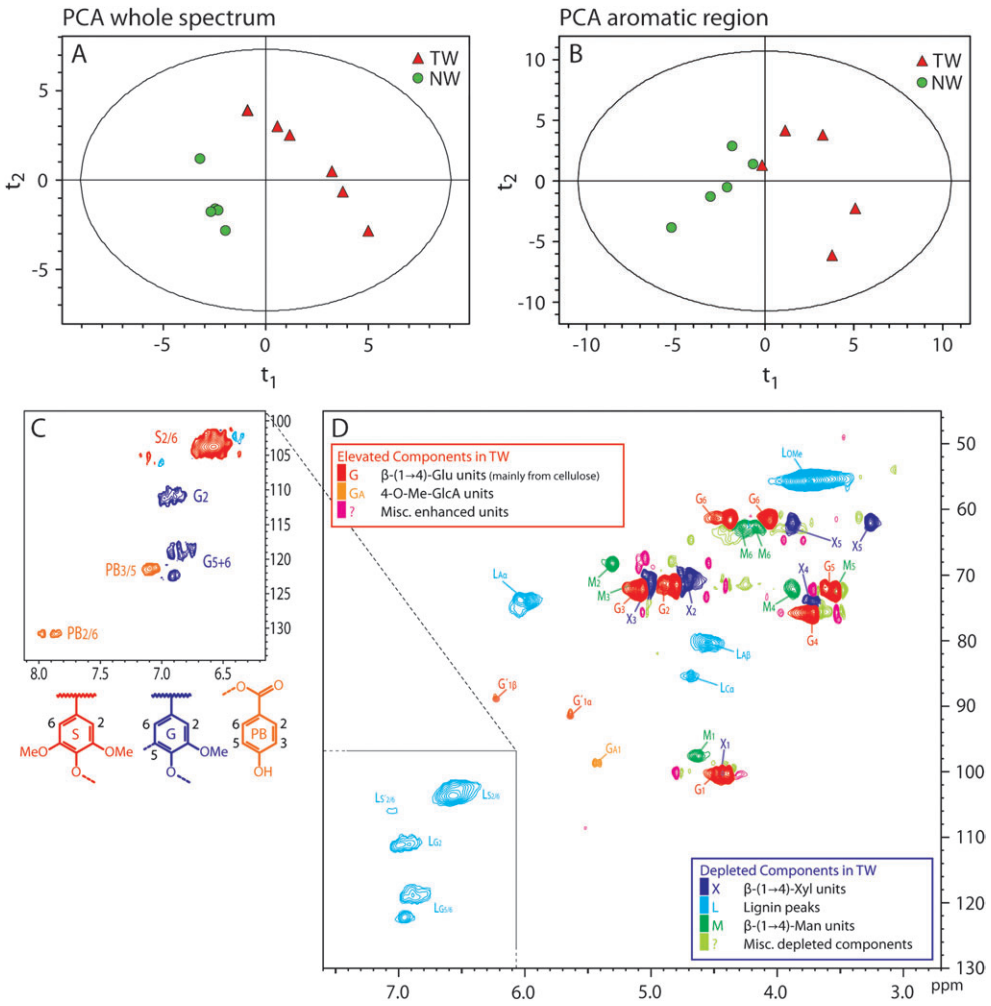

Figure 3. Comparison between Tension Wood and Normal Wood Using PCA.

(A) The $\mathbf{t}_1$–$\mathbf{t}_2$ scores scatter plot from the PCA model using the whole spectral region (with the exception of the acetyl resonances) showing class separation between tension wood and normal wood samples along $\mathbf{t}_1$.

(B) The $\mathbf{t}_1$–$\mathbf{t}_2$ scores scatter plot from the PCA model of the aromatic region of the spectra again showing class separation between tension wood and normal wood samples. One of the TW samples was considered to be an outlier and was removed from this model.

(C) 2D loading plot $\mathbf{p}_1$ explaining the spectral variation along $\mathbf{t}_1$ in the PCA model shown in (B). The red and orange resonances corresponding to S and PB lignin units are elevated and the purple resonances from G lignin units are relatively decreased.

(D) Corresponding 2D loading plot from the PCA model shown in (A). Resonances in red and orange correspond to cell wall components with a relative increase in tension wood compared to normal wood. The resonances in blue and green correspond to components with a relative decrease in tension wood. G′1α and G′1β refer to the α- and β-conformations of the glucosyl anomeric units at the reducing end of cellulose.

One benefit of using full-resolution spectra in the analysis as opposed to peak picking and using a resulting peak list is apparent from the resolution obtained from peaks that overlap with a number of other peaks in the original spectrum (compare the cellulose G6 peaks in Figures 1 and 3D). This is a result of the co-varying behavior of all data points belonging to the same peak. For the cellulose peaks that are responsible for the variation along $\mathbf{t}_1$, their actual 2D-peak shape will be apparent in the loading spectrum, even though they overlap with other peaks in the original spectra. This high-resolution information about peaks in congested regions of the spectra will almost certainly be lost when performing a peak picking analysis.

Changes in lignin composition could not be deduced directly from the PCA model using the full spectral range because of a relative decrease in the total lignin amount. However, the spectral region containing aromatic peaks can be modeled separately to detect changes in lignin composition (S/G-ratio, degree of acylation by *p*-hydroxybenzoate) without interference from polysaccharides. In this case, qualitative differences between the lignin in TW and NW are visualized, ignoring the fact that TW has an overall decrease in lignin. A PCA model using only this spectral region also discriminated between TW and NW (Figure 3B). The loading plot shows that the sample discrimination is a result of the anticipated higher

relative amounts of S lignin and *p*-hydroxybenzoates, and relatively decreased G lignin in tension wood (Figure 3C).

Taken together, the results obtained are consistent with prior knowledge about tension wood chemistry and demonstrate that PCA can be used to detect and, more importantly, interpret differences in cell wall structure between two sample types. In addition, the resulting high-resolution 2D loading plots revealed details that potentially could be lost using peak picking approaches.

Comparison of Cell Wall Composition between PME Down-Regulated and Wild-Type *Populus*

In a second example, we compared wood from transgenic hybrid aspen (*Populus tremula x tremuloides*) down-regulated in PME activity (Siedlecka et al., 2008), with wood from wild-type (WT) trees. The transgenic wood samples in this study were previously shown to be modified in their degree of pectin methylation (Siedlecka et al., 2008). Although pectin chemistry *per se* is difficult to study in the approach used here because a major fraction of the pectins will be extracted during cell wall preparation, a modification in one wall polymer is often observed to affect other components of the wall matrix (Leplé et al., 2007). We use this example to demonstrate the usefulness of combining 2D NMR analysis with multivariate analysis to detect subtle and unpredictable differences between wood samples.

Similar to the previous example, a PCA model from preprocessed 2D ^{13}C–^{1}H HSQC spectra obtained from transgenic and WT wood samples provided an overview of the variation within the data. Unlike the previous comparison between TW and NW, no obvious clustering of sample types was detected (data not shown). It was therefore concluded that eventual differences in the spectra induced by the modification of PME activity are obscured by biological and/or experimental variation. However, the PCA model was still useful as a quality control measure of the input data and variations in processing parameters such as alignment and phase-correction between spectra could be detected and rectified.

In order to investigate whether there exists any systematic variation in the spectral data related to the samples being either transgenics or WT, we used OPLS-DA. In contrast to PCA, this supervised multivariate analysis correlates the variation between spectra with the class identifier (transgenic or WT) of each spectrum. Differences in the spectra correlated to sample type can thus be studied even though the differences account for only a minor part of the total spectral variation within the dataset. To construct a reliable model, the data were scaled to unit variance (UV-scaling). This scaling gives all peaks the same variance, independently of their intensities, which makes it easier to detect differences in cell wall components with low abundance. An additional benefit of this scaling is that the intensities of the peaks in the resulting loading plots are directly correlated to their importance for sample type discrimination. The loadings theoretically have values between ± 1 (corresponding to perfect negative or positive correlation). Simply by choosing an appropriate cut-off value, we can identify peaks with different weights in the model. Although the peak shapes are distorted by the UV-scaling (Craig et al., 2006), the 2D loading plots can still be interpreted based on peak positions (Hedenström et al., 2008).

A two-component OPLS-DA model explained 39% (R^2X 0.39) of the spectral variation between WT and transgenic trees and had a predictive ability (Q^2Y) of 0.34 (Figure 4A). This is not considered to be a strong model, but it still indicates a clear difference between these two sample types. A random permutation test was performed to validate the model and this test confirmed that the model, although rather weak, is reliable. The loading plots (Figure 4B and 4C) were constructed from loadings along $\mathbf{t}_1$ in the score plot (Figure 4A). Positive (red) peaks are thus correlated to transgenic trees and the negative (blue) peaks are correlated to WT trees. It is apparent from Figure 4B that most of the peaks in the spectra to some extent contribute to the sample discrimination. To reveal peaks with the highest correlations to sample discrimination (which are therefore most important for separating the samples), a stringent loading cut-off of 0.5 was applied. This results in a cleaner loading spectrum that facilitates the model interpretation (Figure 4C). Care should be taken, however, when interpreting these loadings because some of the 'peaks' observed in Figure 4C were interpreted as artifacts arising from slight changes in peak positions or line-width for certain peaks between samples rather than a change in peak intensity. Such effects are magnified when using UV-scaled data, as the data points at peak edges have the same variance as all other data points. This effect is obvious for the xylan C5/H5 peaks located at 62/3.9 and 62/3.25 ppm ($^{1}H/^{13}C$ chemical shifts) where only the edges appear to be important, implicating a change in line-width of these peaks. A closer inspection of the individual spectra revealed that these peaks are, in fact, slightly broader for two of the WT samples. Thus, in Figure 4B and 4C, only peaks centered on a peak in the original spectra have been annotated and should be interpreted as increased or decreased levels for a particular cell wall component in the different sample types. Although peak shifts can contain relevant information about changes in cell wall structure, these effects were not further analyzed in this study.

Interpretation of the loading plots (Figure 4B and 4C) revealed that the most prominent differences between sample types appear to be the peaks representing lignin G and *p*-hydroxybenzoate units (Figure 4C), which are negatively correlated with the transgenic samples. To confirm the accuracy of this conclusion, it was verified using derivatization followed by reductive cleavage (the DFRC method) (Lu and Ralph, 1997) and pyrolysis-MS (Supplemental Figure 1). Several peaks originating from different polysaccharides are also seen in the loading spectrum without cut-off (Figure 4B), but only few of these remain in the more stringent loading spectrum (Figure 4C). The cellulose peaks, for example, are prominent in the loading spectrum when no cut-off is used, but are less important to explain the model that discriminates the samples

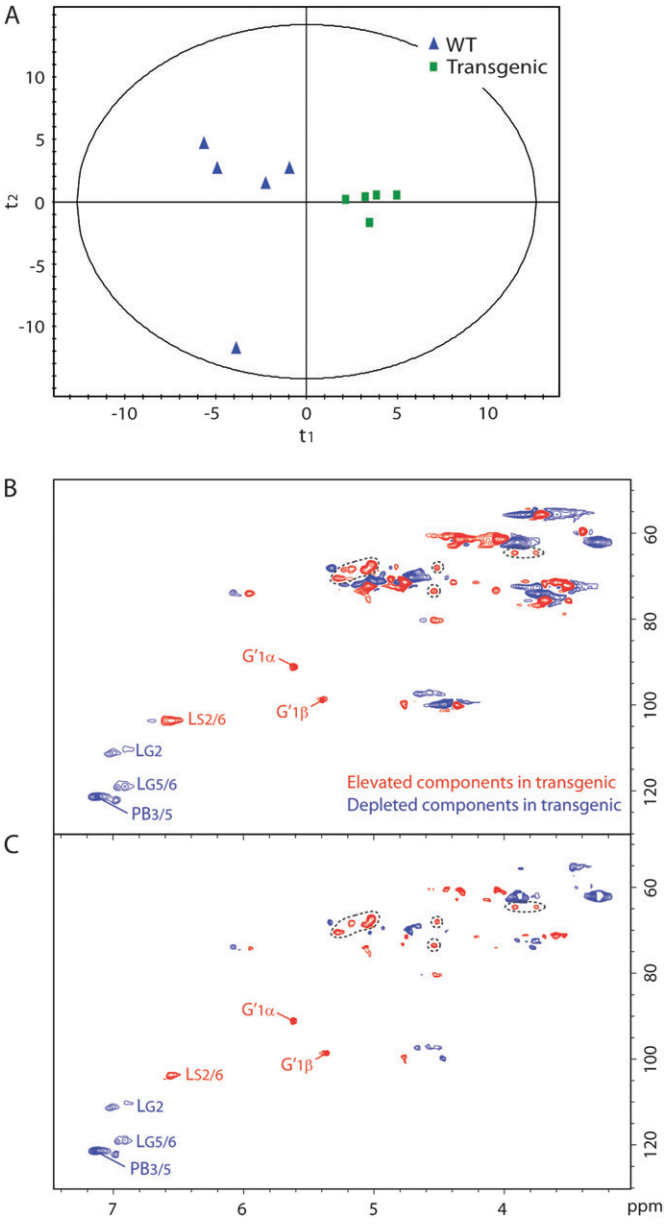

Figure 4. Comparison between Transgenic *Populus* Down-Regulated in Pectin Methyl Esterase Activity with Wild-Type Trees.

(A) The $\mathbf{t}_1$–$\mathbf{t}_2$ scores scatter plot from the OPLS-DA model comparing 2D ^{13}C–^{1}H HSQC spectra from wild-type and transgenic trees. One orthogonal component was calculated in addition to the predictive component and Q^2 was 0.34.

(B) 2D loading spectra constructed from $\mathbf{p}_1$ in the model in Figure 4A showing the differences between the wild-type and the transgenic samples. Red peaks are increased, and blue peaks are decreased, in relative amounts in transgenic trees.

(C) Same loading plots as in (B) but with a cut-off value of 0.5 identifying peaks that are major discriminants between the samples. Annotated peaks are those verified to be either increased or decreased in transgenic samples and the peaks marked with dotted lines are also increased but their identity is as yet unknown. Remaining red and blue peaks are mainly artifacts from subtle peak shifts or variation in line width of peaks between spectra.

as they are filtered away when a 0.5 cut-off was used. Interestingly, some of the unassigned peaks that differed between TW and NW also appear in Figure 4C (marked with dotted lines). It is important to point out that the multivariate approach used here with PCA for detection of eventual outliers and quality control of processing parameters followed by OPLS-DA for sample discrimination results in a simplified and more reliable interpretation of the differences between transgenic plants and WT compared to a traditional difference spectrum (Supplemental Figure 2). This example highlights the usefulness of combining 2D NMR of dissolved samples with a multivariate approach to detect, visualize, and interpret subtle differences in the cell wall chemistry between different sample types.

Concluding Remarks

We have demonstrated that chemometric approaches applied to 2D ^{13}C–^{1}H HSQC NMR spectra can reveal subtle differences between sample types that are not obvious from simple inspection of the spectra. The added ability to interpret the loading spectra as full-resolution 2D spectra provides a powerful new tool to elucidate cell wall structural characteristics. This method is valid for both the lignin and polysaccharide constituents of the cell wall. Additional annotation of polysaccharide peaks in these spectra remains to be performed and will further increase the value of this approach. Although our examples come from secondary walls and wood, the technique should function equally well for samples consisting of primary cell walls. Both intra- and inter-class variation could be interpreted, yielding information about discrimination not only between samples, but also for biological and experimental variation. In the future, we expect the 2D NMR profiles, via these multivariate methods, to be amenable to the same types of correlative analyses that have been used for different types of 1D spectra (Christoffersson et al., 2002; Lestander et al., 2009) to relate cell wall compositional/structural profiles to other variables. Logical applications include predicting biomass processing efficiency and for optimizing pretreatment methods in various biorefinery processes.

METHODS

Plant Materials

TW and NW were collected from ~3-m high 15-year-old field-grown aspen trees (*Populus tremula*) described in Hellgren et al. (2004). TW was induced by fixing the trees at an angle of ~30° with a string at the beginning of the growing season (early June). After the growing season (late August), stem pieces were sampled from midpoint of the stem of bent (TW) and upright (NW) trees, immediately frozen in liquid nitrogen, transported to the lab on dry ice, and stored at –70°. Wood from the latest annual ring (excluding latewood) was used for the analysis. Production and greenhouse culture of transgenic hybrid aspen (*Populus tremula x tremuloides*) trees down-regulated in pectin methylesterase has been described

earlier (Siedlecka et al., 2008) and the samples used here are from the transgenic line 5 in that study. Trees were grown to ~1.5 m, internodes collected from a defined internode at the lower third of the stem, frozen in liquid nitrogen, and stored at –70°.

Sample Preparation and Cell Wall Dissolution

Five samples were prepared for each type of poplar wood investigated, namely tension wood, normal wood, transgenic, and wild-type. Each sample corresponds to a single tree (i.e. no pooling of samples). Removal of extractives (cell wall isolation) from debarked and coarsely ground material was performed by sequential extractions. Cell walls were extracted with water (10 min in an ultrasonic bath), followed by centrifugation (10 min, 4000 rpm) and supernatant removal, and then similarly extracted using 80/20 methanol/water. The methanol/water extraction was performed two additional times. All samples were lyophilized subsequent to cell wall isolation.

Following lyophilization, approximately 400 mg of the wood was ground in a 50-mL ZrO_2 jar with 10 × 10 mm ZrO_2 ball bearings using a Retsch PM100 model planetary ball-mill, for 1 h 50 min using 20-min milling intervals with 10-min breaks in between to avoid excessive heating of the sample. To effect cell wall dissolution (Lu and Ralph, 2003), DMSO (1.8 ml) and *N*-methylimidazole (NMI, 900 μl) were added to 100 mg of each ball-milled wood sample. The samples were stirred on a shaking board for 24 h. Acetylation was accomplished simply by adding acetic anhydride (450 μl) and stirring for 1.5 h. The acetylation reaction was quenched by pouring the acetylated cell wall solution into water (600 ml), where the sample was left to precipitate for 3 h. The samples were then centrifuged in a JA-14 rotor at 21 500 *g* for 10 min. The samples were washed twice following centrifugation under the same conditions. All samples were lyophilized in order to remove remaining solvents.

NMR Acquisition

Approximately 80 mg of prepared acetylated cell wall (Ac-CW) material was dissolved in $CDCl_3$ (600 μl) in a 5-mm NMR tube prior to NMR acquisition. Adiabatic ^{13}C-inversion and ^{13}C-refocusing pulses were implemented in the pulse sequence used in this study (Bruker pulse sequence hsqcetgpsisp.2) (Kay et al., 1992; Kupče and Freeman, 2007; Kupče and Hiller, 2001). We strongly recommend the use of adiabatic pulses in 2D ^{13}C–^{1}H HSQC experiments because of the significant sensitivity improvement and more coherent phase-behavior observed, for instance, for the aromatic resonances. Adiabatic pulses effectively remove off-resonance effects that result in suppressed signal intensities and phase-errors for resonances far from the transmitter frequency. They also remove much of the variation associated with differences in the 1-bond coupling constant, $^{1}J_{C–H}$ (Kupče and Freeman, 1997). The samples were all analyzed with the same optimized acquisition parameters and in randomized order to minimize the effects of experimental variations. All spectra were acquired on a 600-MHz Bruker DRX spectrometer (Bruker Biospin, Rheinstetten, Germany), using a 5-mm TXI cryoprobe equipped with z-gradients. Sine-shaped gradients were used for coherence selection. Sweep widths of 9 and 140 ppm were used in the ^{1}H and ^{13}C dimensions, respectively. The duration of each 2D experiment was approximately 1 h as a result of using a relaxation delay of 1 s and collecting 12 scans for each of the 256 t_1 increments. After zero filling and linear prediction, the resulting data matrix size was 1024 × 512. The spectra were processed using Topspin 1.3 (Bruker Biospin, Rheinstetten, Germany). Gaussian window functions were applied in both dimensions, with LB = –1 and GB = 0.001 in F1 and LB = –0.1 and GB = 0.001 in F2. Baseline correction was performed using a 3rd-order polynomial and all spectra were manually phase corrected and calibrated (using the residual $CHCl_3$ peak at δ_H = 7.27 ppm and δ_C = 77.0 ppm as an internal reference) before analysis.

Multivariate Analysis

Pre-treatment of all 2D ^{13}C–^{1}H HSQC spectra were performed in MATLAB (Mathworks Inc.) using in-house scripts for data import, alignment, selection, and exclusion of spectral regions and visualization of loadings. Multivariate data analysis was performed using SIMCA-P+ 11.0 (MKS-Umetrics, Umeå, www.umetrics.com). The residual $CHCl_3$ resonance was removed from all spectra prior to analysis by setting all data points in this region to zero. Residual resonances from NMI that were present in some spectra from the transgenic plants were also excluded in the same way. The spectral region to be analyzed (whole region except the acetyl region or the aromatic region) was unfolded into a row vector (as shown in Figure 2) that was normalized to a constant sum in order to eliminate differences in spectral intensities related to variations in sample amount. Data points with intensity lower than a noise threshold (determined from a 'noise region' of the spectra) were excluded. The effect of different scaling, mean centering (i.e. no scaling), pareto scaling, and scaling to unit variance (UV-scaling) was evaluated in both studies and the best models were found when using mean centered data in the study of tension wood and UV-scaling for the OPLS-DA in the study of transgenic *Populus* samples. The resulting data matrix consisting of the spectral data of each sample as a row vector was subsequently imported in SIMCA-P+ 11.0 for multivariate analysis. The loading vectors extracted from the models were transferred to MATLAB and the loading values were placed in their original position in a vector with length corresponding to the unfolded original spectrum. Folding of this vector into a matrix with the dimensions 512 × 1024 (as for the original spectra) resulted in the 2D loading spectra. MATLAB scripts for pre-treatment of 2D NMR spectra and visualization of loading are available from the authors upon request.

SUPPLEMENTARY DATA

Supplementary Data are available at *Molecular Plant Online.*

FUNDING

This work was supported by Funcfiber, a FORMAS center for excellence in wood science, EU grant (028974) (CASPIC) and the Swedish energy agency. The US group was supported by the Office of Science (BER), US Dept. of Energy, Interagency agreement No. DE-AI02-06ER64299, and was also funded in part by the DOE Great Lakes Bioenergy Research Center (www.greatlakesbioenergy.org), which is supported by the US Department of Energy, Office of Science, Office of Biological and Environmental Research, through Cooperative Agreement DE-FC02-07ER64494 between The Board of Regents of the University of Wisconsin System and the US Department of Energy.

ACKNOWLEDGMENTS

We are grateful to Daniel Yelle (US Forest Products Lab, and U. Wisconsin Dept. of Forestry) for help with polysaccharide assignments. No conflict of interest declared.

REFERENCES

Boerjan, W., Ralph, J., and Baucher, M. (2003). Lignin biosynthesis. Annu. Rev. Plant Biol. **54,** 519–549.

Bowling, A., and Vaughn, K. (2008). Immunocytochemical characterization of tension wood: gelatinous fibers contain more than just cellulose. Am. J. Bot. **95,** 655–663.

Christoffersson, K.E., Sjöström, M., Edlund, U., Lindgren, A., and Dolk, M. (2002). Reactivity of dissolving pulp: characterisation using chemical properties, NMR spectroscopy and multivariate data analysis. Cellulose. **9,** 159–170.

Craig, A., Cloarec, O., Holmes, E., Nicholson, J.K., and Lindon, J.C. (2006). Scaling and normalization effects in NMR spectroscopic metabonomic data sets. Anal. Chem. **78,** 2262–2267.

Gierlinger, N., and Schwanninger, M. (2006). Chemical imaging of poplar wood cell walls by confocal Raman microscopy. Plant Physiol. **140,** 1246–1254.

Hedenström, M., Wiklund, S., Sundberg, B., and Edlund, U. (2008). Visualization and interpretation of OPLS models based on 2D NMR data. Chemom. Intell. Lab. Sys. **92,** 110–117.

Hellgren, J.M., Olofsson, K., and Sundberg, B. (2004). Patterns of auxin distribution during gravitational induction of reaction wood in poplar and pine. Plant Physiol. **135,** 212–220.

Holmes, E., et al. (2000). Chemometric models for toxicity classification based on NMR spectra of biofluids. Chem. Res. Toxicol. **13,** 471–478.

Hu, Z., Yeh, T.-F., Chang, H.-m., Matsumoto, Y., and Kadla, J.F. (2006). Elucidation of the structure of cellulolytic enzyme lignin. Holzforschung. **60,** 389–397.

Ikeda, T., Holtman, K., Kadla, J.F., Chang, H.M., and Jameel, H. (2002). Studies on the effect of ball milling on lignin structure using a modified DFRC method. J. Agr. Food Chem. **50,** 129–135.

Jackson, J.E. (1991). A Users Guide to Principal Components (New York: John Wiley).

Kay, L.E., Keifer, P., and Saarinen, T. (1992). Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. J. Am. Chem. Soc. **114,** 10663–10665.

Kim, H., Ralph, J., and Akiyama, T. (2008). Solution-state 2D NMR of ball-milled plant cell wall gels in DMSO-d_6. BioEnergy Res. **1,** 56–66.

Kupče, E., and Freeman, R. (1997). Compensation for spin–spin coupling effects during adiabatic pulses. J. Magn. Reson. **127,** 36–48.

Kupče, E., and Freeman, R. (2007). Compensated adiabatic inversion pulses: broadband INEPT and HSQC.J. Magn. Reson. **187,** 258–265.

Kupče, E., and Hiller, W. (2001). Clean adiabatic TOCSYs. Magn. Reson. Chem. **39,** 231–235.

Leplé, J.-C., et al. (2007). Downregulation of cinnamoyl-coenzyme A reductase in poplar: multiple-level phenotyping reveals effects on cell wall polymer metabolism and structure. Plant Cell. **19,** 3669–3691.

Lestander, T.A., Johnsson, B., and Grothage, M. (2009). NIR techniques create added values for the pellet and biofuel industry. Bioresource Technology. **100,** 1589–1594.

Lu, F.C., and Ralph, J. (1997). DFRC method for lignin analysis. 1. New method for beta aryl ether cleavage: lignin model studies. J. Agr. Food Chem. **45,** 4655–4660.

Lu, F.C., and Ralph, J. (2003). Non-degradative dissolution and acetylation of ball-milled plant cell walls: high-resolution solution-state NMR. Plant J. **35,** 535–544.

Maunu, S.L. (2002). NMR studies of wood and wood products. Prog. Nucl. Magn. Reson. Spectrosc. **40,** 151–174.

O'sullivan, A.C. (1997). Cellulose: the structure slowly unravels. Cellulose. **4,** 173–207.

Pilate, G., et al. (2004). Lignification and tension wood. C.R. Biol. **327,** 889–901.

Ralph, J., and Lu, F.C. (2004). Cryoprobe 3D NMR of acetylated ball-milled pine cell walls. Org. Biomol. Chem. **2,** 2714–2715.

Ralph, J., et al. (1999). Solution-state NMR of lignins. In: Advances in Lignocellulosics Characterization, Argyropoulos, D.S., ed. (Atlanta, GA: TAPPI Press), pp. 55–108.

Ralph, J., et al. (2006). Effects of coumarate-3-hydroxylase downregulation on lignin structure. J. Biol. Chem. **281,** 8843–8853.

Ralph, J., et al. (2008). Identification of the structure and origin of a thioacidolysis marker compound for ferulic acid incorporation into angiosperm lignins (and an indicator for cinnamoyl CoA reductase deficiency). Plant J. **53,** 368–379.

Ralph, S.A., Landucci, L.L., and Ralph, J. (2005). NMR database of lignin and cell wall model compounds. available online at http://ars.usda.gov/Services/docs.htm?docid=10429 (previously www.dfrc.ars.usda.gov/software.html), updated at least annually since 1993.

Siedlecka, A., et al. (2008). Pectin methyl esterase inhibits intrusive and symplastic cell growth in developing wood cells of *Populus*. Plant Physiol. **146,** 554–565.

Stenlund, H., Gorzsas, A., Persson, P., Sundberg, B., and Trygg, J. (2008). Orthogonal projections to latent structures discriminant analysis modeling on in situ FT-IR spectral imaging of liver tissue for identifying sources of variability. Anal. Chem. **80,** 6898–6906.

Timell, T.E. (1969). Chemical composition of tension wood. Svensk Papperstidn. **72,** 173–181.

Trygg, J., and Wold, S. (2002). Orthogonal projections to latent structures (O-PLS). J. Chemometr. **16,** 119–128.

Tsuchikawa, S. (2007). A review of recent near infrared research for wood and paper. Appl. Spectrosc. Rev. **42,** 43–71.

Wagner, A., et al. (2007). Modifying lignin in conifers: the role of HCT during tracheary element formation in *Pinus radiata*. Proc. Natl Acad. Sci. U S A. **104,** 11856–11861.

Wold, S., Antti, H., Lindgren, F., and Öhman, J. (1997). Orthogonal Signal Correction of Near-Infrared Spectra, 5th Scandinavian Symposium on Chemometrics (SSC5) (Lathi, Finland, Elsevier Science Bv.), pp. 175–185.

Wold, S., Esbensen, K., and Geladi, P. (1987). Principal component analysis. Chemom. Intell. Lab. Sys. **2,** 37–52.

Yelle, D.J., Ralph, J., and Frihart, C.R. (2008). Characterization of non-derivatized plant cell walls using high-resolution solution-state NMR spectroscopy. Magn. Reson. Chem. **46,** 508–517.

Yoshida, M., Ohta, H., Yamamoto, H., and Okuyama, T. (2002). Tensile growth stress and lignin distribution in the cell walls of yellow poplar, *Liriodendron tulipifera* Linn. Trees-Struct. Funct. **16,** 457–464.